Date: Dec. 28 2018

EIC Detector R&D Progress Report

Project ID: eRD20
Project Name: Developing Simulation and Analysis Tools for the EIC
Period Reported: Oct. 1 2018 - Dec. 28 2018
Project Leaders: M. Diefenthaler, A. Kiselev
Contact Persons: M. Diefenthaler (mdiefent@jlab.org), A. Kiselev (ayk@bnl.gov)

Project members

- **ANL** Whitney Armstrong, Sergei Chekanov, David Blyth
- **BNL** Elke-Caroline Aschenauer, Alexander Kiselev, Jerome Lauret, Christopher Pinkenburg
- **INFN Trieste** Andrea Bressan
- **Jefferson Lab** Markus Diefenthaler, David Lawrence, Dmitry Romanov, Maurizio Ungaro
- **Lund** Stefan Prestel
- **SLAC** Makoto Asai, Dennis Wright
- **William & Mary** Wouter Deconinck

Abstract

Developing the physics program for the EIC, and designing the detectors needed to realize it, requires a plethora of software tools and multifaceted analysis efforts. Many of these tools have yet to be developed or need to be expanded and tuned for the physics reach of the EIC. Currently, various groups use disparate sets of software tools to achieve the same or similar analysis tasks such as Monte Carlo event generation, interaction region and detector simulations, track reconstruction, and event visualization to name a few examples. With a long-range goal of the successful execution of the EIC scientific program in mind, it is clear that early investment in the development of well-defined interfaces for communicating, sharing, and collaborating, will facilitate a timely completion of not just the planning and design of an EIC but ultimate delivery of the physics capable with an EIC.

Past

What was planned for this period?

In our proposal, we have identified five major goals for FY19:

- ❏ Development of Detector Simulations with a focus on:
  - ❏ Geant4 Simulations (proposal p. 5)
  - ❏ Interfaces and integration, with priority on EIC Software Sandbox and Community Reference Reconstruction (proposal p. 6)
- ❏ Initiatives for:
  - ❏ Containers (proposal p. 6)
  - ❏ Artificial Intelligence (AI) and Machine Learning (ML) (proposal p. 8)
  - ❏ Monte Carlo Event Generators (proposal p. 9)

What was achieved?

We are the core of the EIC User Group Software Working Group (ESWG) and have started to collaborate in FY19 with the HEP Software Foundation (HSF) and the ROOT team at CERN:

- ❏ There will be joint ESC/ESWG-HSF meetings and common work on artificial intelligence, event reconstruction, and real-time analysis.
- ❏ Enric Saavedra (CERN) will be the liaison between the ESC/ESWG and the ROOT team. We will work on a programming models for the EIC to significantly improve the usability of the EIC software and on parallelism on future computing environments.

On November 6, we organized a ML seminar to review promising R&D areas in ML in HEP and NP. Anima Anandkumar (Caltech, Nvidia) addressed in her keynote “Advances in Trinity of AI: Data, Algorithms & Compute”. An overview about the R&D on scientific machine learning was provided by Steven Lee (DOE ASCR) and Wahid Bhimji (NERSC). A.-K. Burns (William & Mary), C. Fanelli (MIT), and A. Tsaris (Fermilab) presented specific use cases for ML methods in HEP and NP.

In cooperation with the MCnet we organize in February the second workshop on “**MCEGs for future ep and eA facilities**”. The workshop will take place at DESY Hamburg on February 20 – 22 2019. We will review recent R&D on Monte Carlo Event Generators (MCEGs) for lepton-proton (ep) and lepton-nucleus (eA) collisions and will discuss the requirements and R&D needs for the MCEGs for the future EIC, LHeC, and VHEeP facilities.

We are involved in the organization of the “**Joint HEP Software Foundation Workshop, Open Science Grid All-Hands Meeting,** and the **Worldwide**

LHC Computing Grid Workshop“ on March 18-21 at Jefferson Lab and will present on ESC/ESWG and EIC Software in general in the plenary sessions.

We help organizing the “**Geant4 Collaboration Meeting**“ which will be held on September 16-20 at Jefferson Lab. The meeting will include a technical forum on EIC detectors.

EIC Software Sandbox and Community Reference Reconstruction

We are setting up a “software sandbox” environment, which is supposed to unify to a large extent presently disconnected event simulation and reconstruction pieces of code used by our various communities (at ANL, BNL and JLAB in particular). This unified EIC Software Sandbox environment should greatly improve the coherence of our efforts in physics simulations, detector modeling as well as the machine IR background studies, in particular in terms of comparing the performance of the proposed accelerator and EIC detector configurations. The working proposal, which is based on the noticeable progress we made over the last two years in terms of defining EIC event I/O model, interchange file formats and container infrastructure was finally shaped up during the recent eRD20 meeting at William & Mary in May 2018. The sandbox will represent itself as a Docker and singularity container image with a well-defined set of “checkpoints”, namely the I/O standards describing:

1) Monte-Carlo input events,
2) digitized hits after transporting these events through a particular detector geometry in a particular software framework and/or toolkit,
3) reconstructed physics events as output.

It is anticipated that any EIC detector simulation which we have available at the moment (TOPSiDE at ANL, EicRoot and fun4all at BNL, GEMC at JLAB) should be able to perform a “transition” step from the checkpoint #1 to the checkpoint #2 according to its own way of implementing the particle transport and generating hits in the sensitive detectors.

We have started with the tracking implementation, the EIC Community Reference Reconstruction, and use JANA (a mature and well-maintained piece of JLAB Hall D in-house software) as a workflow tool. Once the detector hits are represented in a unified way at the checkpoint #2, it is assumed that a very generic package based on a straightforward GENFIT implementation will do the rest of the track reconstruction independently of the internal details of a particular software framework which produced the file with the digitized hits and will represent the output events in a unified standard at checkpoint #3. This way the internal machinery of a particular software framework is hidden completely from the end user, who is typically only interested in comparing of reconstructed events to the simulated ones while (as mentioned earlier) both will be available in a framework-independent way. We believe that since in this approach the digitized hit information is

effectively decoupled from the geometry details (material distribution in particular), it should be sufficient for the track reconstruction code to be able to import just a standard GDML or ROOT TGeo file and a magnetic field map of a given setup in order to perform the track fitting. The unified format for the magnetic field maps needs to be worked out, but does not look like a task of overwhelming complexity to us. Track finding can be implemented at a later stage. The singularity flavor of this sandbox container image should allow easy portability of the modeling infrastructure between the desktop environment and the production grade computing resources.

D.Romanov with D. Lawrence started the implementation of the Community Reference Reconstruction. The software is based on the JANA workflow tool. The source code is available at https://gitlab.com/eic/ejana (the name “ejana” comes from EIC JANA). One of the strong advantages of the framework is a plugin system which allows to achieve several important goals:

- First, hiding the internal complexity, it allows to automatically activate different parts and utilize various data processing paths depending on execution context.
- Which in turns allows to build easily extensible software that could be used in different scenarios, analysis, testing, calibration, etc., and environments, ordinary PCs, computing farms, cloud services or even real time. The latter is important in the context of the R&D by the EIC Streaming Readout Consortium.
- The plugin system also helps users to organize a workflow in such a way, that independent groups of experts could work in parallel with minimal code interference with each other during development cycles.

The first goal to achieve was to get GENFIT tracking and RAVE vertexing to work in terms of JANA framework (and its plugins), utilizing standard ROOT TGeo or GDML (through TGeo) formats for geometry input, TTree and TFile for data input/output and a magnetic field given in a text file. The overall architecture and plugin development takes into account the final goal to have the framework which is easily extendable and is agnostic to source of data input.

At this point the framework major dependencies are ROOT, JANA, GENFIT and RAVE. The build system takes into account that each of dependent packages may be optionally excluded in the future. In reference to EIC Container Initiative, eJANA is to be released on several docker images: with minimal dependencies, on top of JLAB environment and later on top of other community images.

D. Romanov introduced “EJPM” packet which stands for EIC JANA Packet Manager. Located https://gitlab.com/eic/ejpm . The software main goal is to unify eJANA application deployment workflow in different environments: containers (in reference to Container Initiative), ordinary user machines and

production (computing) environments. Secondary goal, is to enhance user experience of working on eJANA (and its plugins) development cycles.

D. Romanov coordinated within the PID consortium (eRD14) a series of meetings about the current status of reconstruction software of PID subsystems and future possibilities of integration into the Community Reference Reconstruction. A detailed review of each PID subsystem has been done and a plan for the integration has been proposed.

What was not achieved, why not, and what will be done to correct?

Apart from the workshop organization, we have put our sole focus in FY19 on the community reference reconstruction which will remain our focus until the end of the second quarter. The work on all other tasks will resume in the third and last quarter of FY19.

Our FY19 proposal included a request for "EIC Software tour" events across the U.S. to popularize the EIC software tools among the community. While we have not received full funding, we are planning an EIC Software tour for late spring or early summer of 2019 and are preparing in-person tutorials for the EIC simulation and reconstruction tools. We will revise our tutorials and documentation according to the experience we will gain in the first events. Based on this experience, we will also develop online tutorials.

Future

We will report on the plans for the next funding cycle and beyond in the progress report for the July meeting.

Manpower

The eRD20 members work on a best effort basis on our projects. The eRD20 funds are presently not used to pay any positions, including the two PostDocs on our project.

External Funding

No external funding has been obtained.

Publications

We have not yet published any reports in FY19.